# Jiatong Shi

*Gates and Hillman Centers, 4902 Forbes Ave, Pittsburgh, PA 15213*
jiatongs@cs.cmu.edu | shijt.site | ftshijt

## Education

**Carnegie Mellon University** — *Pittsburgh, PA, U.S.A.*
PhD in Language Technologies — *2021.08 - 2026.02 (Expected)*
- Advisor: Dr. Shinji Watanabe

**The Johns Hopkins University** — *Baltimore, MD, U.S.A.*
MS Computer Science — *2019.08 - 2021.05*
- Advisor: Dr. Shinji Watanabe

**Renmin University of China** — *Beijing, China*
BS Computer Science — *2015.09 - 2019.06*
- Second Major in Fintech
- Undergrad research advisor: Dr. Qin Jin, Dr. Wei Xu, and Dr. Wei Du

## Research Experience

**Carnegie Mellon University - Language Technologies Institute** — *Pittsburgh, PA, U.S.A.*
Advisor: Dr. Shinji Watanabe — *2021 - present*
- Focus: speech representation learning

**Frederick Jelinek Memorial Summer Workshop (JSALT, JHU Summer Workshop)** — *Baltimore, MD, U.S.A.*
Advisor: Dr. Ann Lee, Dr. Shinji Watanabe, and Dr. Hung-yi Lee — *2022 Summer*
- Focus: unsupervised ASR and its applications

**The Johns Hopkins University - Center for Language and Speech Processing (CLSP)** — *Baltimore, MD, U.S.A.*
Advisor: Dr. Shinji Watanabe — *2019 - 2021*
- Focus: low-resource, multilingual, robust speech recognition and applications in endangered language documentation

**Renmin University of China - Multi-level Multi-aspect Multimedia Analysis (AI.M3) Lab** — *Beijing, China*
Advisor: Dr. Qin Jin — *2018-2019*
- Thesis: “Computer assisted language learning for young English learners”
- Focus: computer assisted language learning (CALL) and music processing

**Renmin University of China - Lab for Information System Applications (LISA)** — *Beijing China*
Co-Advisors: Dr. Wei Xu and Dr. Wei Du — *2017-2019*
- Thesis: “Multimedia Analytics for Crowdfunding Prediction”
- Focus: online data analytic and fin-tech applications

## Professional Experience

**Anuttacon** — *Mountain View, CA, U.S.A.*
Co-Supervisors: Cong Zhou and Santiago Pascual — *May. 2025 - Present*
- Project: Speech evaluation foundation model with a focus on voice role-play evaluation.

**Tencent America (Tencent AI Lab)** — *Bellevue, WA, U.S.A.*
Co-Supervisors: Dr. Chunlei Zhang and Dr. Dong Yu — *May. 2024 - Aug. 2024*
- Project: Speech language models for speech processing.

**Meta AI** — *Menlo Park, CA, U.S.A.*
Co-Supervisors: Anna Sun, Dr. Xutai Ma, Dr. Hirofumi Inaguma, and Dr. Ilia Kulikov — *May. 2023 - Dec. 2023*
- Project: Self-supervised speech representation learning and its application to speech-to-speech translation.

**Tencent America (Tencent AI Lab)** *Bellevue, WA, U.S.A.*

Co-Supervisors: Dr. Chunlei Zhang and Dr. Dong Yu *May. 2022 - Aug. 2022*

- Project: Towards streaming speaker diarization systems in noisy meeting scenarios.

**IBM Watson Research Center** *Yorktown Heights, NY, U.S.A.*

Supervisor: Dr. George Saon *May. 2021 - Aug. 2021*

- Project: Lattice generation, confusion network processing, and confidence measure estimation for RNN transducer

**Tencent America (Tencent AI Lab)** *Bellevue, WA, U.S.A.*

Co-Supervisors: Dr. Chunlei Zhang, Dr. Chao Weng, and Dr. Dong Yu *May. 2020 - Dec. 2020*

- Project 1: the joint-framework for target speaker extraction and speech recognition
- Project 2: end-to-end speaker diarization with generalized neural speaker clustering

**Tencent Ethereal Audio Lab** *Shenzhen, Guangdong, China*

Supervisor: Wei (Dennis) Xiao *Jun. 2019 – Aug. 2019*

- Project 1: psycho-acoustic loss design for speech super-resolution
- Project 2: real-time speech super-resolution system for Tencent Meeting (i.e., Voov Meeting)

**NetEase Youdao - AI Application** *Beijing, China*

Supervisor: Yixuan Xiao *May. 2018 - Nov. 2018*

- Project: a prototype system for computer assisted language learning

## Publications

** equal contribution; + mentored students;*

### Published

99. **Jiatong Shi**, Yifan Cheng+, Bo-Hao Su, Hye-jin Shim, Jinchuan Tian, Samuele Cornell, Yiwen Zhao+, Siddhant Arora, and Shinji Watanabe. 2025. "ARECHO: Autoregressive Evaluation via Chain-Based Hypothesis Optimization for Speech Multi-Metric Estimation" (Accepted by NeurIPS2025 Spotlight)

98. Yiwen Zhao+, **Jiatong Shi**, Jinchuan Tian, Yuxun Tang+, Jiarui Hai, Jionghao Han+, and Shinji Watanabe. 2025. "Adapting Speech Language Model to Singing Voice Synthesis" (Accepted by NeurIPS AI4Music Workshop)

97. Pooneh Mousavi, Gallil Maimon*, Adel Moumen*, Darius Petermann*, **Jiatong Shi***, Haibin Wu*, Haici Yang*, Anastasia Kuznetsova*, Artem Ploujnikov, Ricard Marxer, Bhuvana Ramabhadran, Benjamin Elizalde, Loren Lugosch, Jinyu Li, Cem Subakan, Phil Woodland, Minje Kim, Hung-yi Lee, Shinji Watanabe, Yossi Adi, Mirco Ravanelli. 2025. "Discrete Audio Tokens: More Than a Survey!" (Accepted by TMLR)

96. **Jiatong Shi**, Chunlei Zhang, Jinchuan Tian, Junrui Ni, Hao Zhang, Shinji Watanabe, and Dong Yu. 2025. "Balancing Speech Understanding and Generation: an Investigation of Continual Pre-training for Codec-based Speech LLM on Translation Tasks" (Accepted by ASRU2025)

95. **Jiatong Shi**, Bo-Hao Su, Shikhar Bharadwaj, Yiwen Zhao+, Shih-Heng Wang, Jionghao Han+, Haoran Wang+, Wei Wang, Wenhao Feng+, Yuxun Tang+, Nezih Topaloğlu, Siddhant Arora, Jinchuan Tian, William Chen, Hye-jin Shim, Wangyou Zhang, Wen-Chin Huang, and Shinji Watanabe. 2025. "VERSA-v2: A Modular and Scalable Toolkit for Speech and Audio Evaluation with Expanded Metrics, Visualization, and LLM Integration" (Accepted by ASRU2025)

94. **Jiatong Shi**, Haoran Wang+, William Chen, Chenda Li, Wangyou Zhang, Jinchuan Tian, and Shinji Watanabe. 2025. "PURE Codec: Progressive Unfolding of Residual Entropy for Speech Codec Learning" (Accepted by ASRU2025)

93. Wei Wang, Wangyou Zhang, Chenda Li, **Jiatong Shi**, Shinji Watanabe, and Yanmin Qian. 2025. "Improving Speech Enhancement with Multi-Metric Supervision from Learned Quality Assessment" (Accepted by ASRU2025)

92. Yiwen Zhao+, Jiatong Shi, Yuxun Tang+, William Chen, and Shinji Watnabe. 2025. "Robust Training of Singing Voice Synthesis Using Prior and Posterior Uncertainty" (Accepted by ASRU2025)

91. Yichen Huang, Zachary Novack, Koichi Saito, **Jiatong Shi**, Shinji Watanabe, Yuki Mitsufuji, John Thickstun, and Chris Donahue. 2025. "Aligning Text-to-Music Evaluation with Human Preferences" (Accepted by ISMIR2025)

90. Yifan Cheng+, Ruoyi Zhang, and **Jiatong Shi**. "MIKU-PAL: An Automated and Standardized Multi-Modal Method for Speech Paralinguistic and Affect Labeling". *Proceedings of Interspeech*, pp. 4308–4312.

89. Zaid Sheikh, Shuichiro Shimizu, Siddhant Arora, **Jiatong Shi**, Samuele Cornell, Xinjian Li, and Shinji Watanabe. "Scalable Spontaneous Speech Dataset (SSSD): Crowdsourcing Data Collection to Promote Dialogue Research". *Proceedings of Interspeech*, pp. 3963–3967.

88. Siddhant Arora, Jinchuan Tian, Hayato Futami, Jee-weon Jung, **Jiatong Shi**, Yosuke Kashiwagi, Emiru Tsunoo, and Shinji Watanabe. "Chain-of-Thought Training for Open E2E Spoken Dialogue Systems". *Proceedings of Interspeech*, pp. 4833–4837.

87. **Jiatong Shi**, Hye-jin Shim, and Shinji Watanabe. "Uni-VERSA: Versatile Evaluation of Speech with a Unified Framework" *Proceedings of Interspeech*, pp. 1798–1802.

86. Jinchuan Tian, William Chen, Yifan Peng, **Jiatong Shi**, Siddhant Arora, Shikhar Bharadwaj, Maekaku Takashi, Yusuke Shinohara, Keita Goto, Xiang Yue, Chao-Han Huck Yang, and Shinji Watanabe. "OpusLM: A Family of Open Unified Speech Language Models". *Proceedings of Interspeech*, pp. 3259–3263.

85. William Chen, Chutong Meng, **Jiatong Shi**, Martijn Bartelds, Shih-Heng Wang[+], Hsiu-Hsuan Wang, Rafael Mosquera, Sara Hincapie, Dan Jurafsky, Antonis Anastasopoulos, Hung-yi Lee, Karen Livescu, and Shinji Watanabe. "The ML-SUPERB 2.0 Challenge: Towards Inclusive ASR Benchmarking for All Language Varieties". *Proceedings of Interspeech*, pp. 2093–2097.

84. Mingda Liu[+], and **Jiatong Shi**. "Bridging Speech and Singing: Multi-stage Speech-Prompted Singing Voice Conversion with Speaker Embedding Adaptation". *Proceedings of Interspeech*, pp. 1248–1252.

83. **Jiatong Shi**, Hye-jin Shim, Jinchuan Tian, Siddhant Arora, Haibin Wu, Darius Petermann, Jia Qi Yip, You Zhang, Yuxun Tang[+], Wangyou Zhang, Dareen Safar Alharthi, Yichen Huang, Koichi Saito, Jionghao Han[+], Yiwen Zhao[+], Chris Donahue, and Shinji Watanabe. 2025. "VERSA: A Versatile Evaluation Toolkit for Speech, Audio, and Music". *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (System Demonstrations)*, pp. 191–209.

82. Siddhant Arora, Yifan Peng, **Jiatong Shi**, Jinchuan Tian, William Chen, Shikhar Bharadwaj, Hayato Futami, Yosuke Kashiwagi, Emiru Tsunoo, Shuichiro Shimizu, Vaibhav Srivastav, and Shinji Watanabe. 2025. "ESPnet-SDS: Unified Toolkit and Demo for Spoken Dialogue Systems". *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (System Demonstrations)*, pp. 248–259.

81. Jinchuan Tian, **Jiatong Shi**, William Chen, Siddhant Arora, Yoshiki Masuyama, Takashi Maekaku, Yihan Wu, Junyi Peng, Shikhar Bharadwaj, Yiwen Zhao[+], Samuele Cornell, Yifan Peng, Xiang Yue, Chao-Han Huck Yang, Graham Neubig, and Shinji Watanabe. 2025. "ESPnet-SpeechLM: An Open Speech Language Model Toolkit". *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (System Demonstrations)*, pp. 116–124.

80. Chien-yu Huang, Wei-Chih Chen, Shu-wen Yang, Andy T. Liu, Chen-An Li, Yu-Xiang Lin, Wei-Cheng Tseng, Anuj Diwan, Yi-Jen Shih, **Jiatong Shi**, William Chen, Xuanjun Chen, Chi-Yuan Hsiao, Puyuan Peng, Shih-Heng Wang, Chun-Yi Kuan, Ke-Han Lu, Kai-Wei Chang, Chih-Kai Yang, Fabian Alejandro Ritter Gutierrez, Huang Kuan-Po, Siddhant Arora, You-Kuan Lin, Chuang Ming To, Eunjung Yeo, Kalvin Chang, Chung-Ming Chien, Kwanghee Choi, Cheng-Hsiu Hsieh, Yi-Cheng Lin, Chee-En Yu, I-Hsiang Chiu, Heitor Guimarães, Jionghao Han[+], Tzu-Quan Lin, Tzu-Yuan Lin, Homu Chang, Ting-Wu Chang, Chun Wei Chen, Shou-Jen Chen, Yu-Hua Chen, Hsi-Chun Cheng, Kunal Dhawan, Jia-Lin Fang, Shi-Xin Fang, Kuan Yu Fang Chiang, Chi An Fu, Hsien-Fu Hsiao, Ching Yu Hsu, Shao-Syuan Huang, Lee Chen Wei, Hsi-Che Lin, Hsuan-Hao Lin, Hsuan-Ting Lin, Jian-Ren Lin, Ting-Chun Liu, Li-Chun Lu, Tsung-Min Pai, Ankita Pasad, Shih-Yun Shan Kuan, Suwon Shon, Yuxun Tang[+], Yun-Shao Tsai, Wei Jui Chiang, Tzu-Chieh Wei, Chengxi Wu, Dien-Ruei Wu, Chao-Han Huck Yang, Chieh-Chi Yang, Jia Qi Yip, Shao-Xiang Yuan, Haibin Wu, Karen Livescu, David Harwath, Shinji Watanabe, and Hung-yi Lee. 2025. "Dynamic-SUPERB Phase-2: A Collaboratively Expanding Benchmark for Measuring the Capabilities of Spoken Language Models with 180 Tasks". *Proceedings of the Twelfth International Conference on Learning Representations (ICLR)*.

79. Yui Sudo, Shakeel Muhammad, Yosuke Fukumoto, Brian Yan, **Jiatong Shi**, Yifan Peng, and Shinji Watanabe. 2025. "Joint Beam Search Integrating CTC, Attention, and Transducer Decoders". *IEEE Transactions on Audio, Speech and Language Processing*, vol. 33, pp. 598-612

78. Jinchuan Tian, Chunlei Zhang, **Jiatong Shi**, Hao Zhang, Jianwei Yu, Shinji Watanabe, and Dong Yu. 2025. "Preference Alignment Improves Language Model-Based TTS". *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1-5.

77. William Chen, Wangyou Zhang, Yifan Peng, Xinjian Li, **Jiatong Shi**, Jinchuan Tian, Xuankai Chang, Soumi Maiti, Karen Livescu, and Shinji Watanabe. 2024. "Towards Robust Speech Representation Learning for Thousands of Languages". *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 10205–10224.

76. **Jiatong Shi**[*], Jinchuan Tian[*], Yihan Wu[*], Jee-weon Jung, Jia Qi Yip, Yoshiki Masuyama, William Chen, Yuning Wu[+], Yuxun Tang[+], Massa Baali, Dareen Alharthi, Dong Zhang, Ruifan Deng, Tejes Srivastava[+], Haibin Wu, Alexander Liu, Bhiksha Raj, Qin Jin, Ruihua Song, and Shinji Watanabe. 2024. "ESPnet-Codec: Comprehensive Training and Evaluation of Neural Codecs for Audio, Music, and Speech". *Proceedings of IEEE Spoken Language Technology Workshop (SLT)*, pp. 562-569.

75. You Zhang, Yongyi Zang, **Jiatong Shi**, Ryuichi Yamamoto, Tomoki Toda, and Zhiyao Duan. 2024. "SVDD 2024: The Inaugural Singing Voice Deepfake Detection Challenge". *Proceedings of IEEE Spoken Language Technology Workshop (SLT)*, pp. 782-787.

74. Masao Someki, Kwanghee Choi, Siddhant Arora, William Chen, Samuele Cornell, Jionghao Han[+], Yifan Peng, **Jiatong Shi**, Vaibhav Srivastav, and Shinji Watanabe. 2024. "ESPnet-EZ: Python-only ESPnet for Easy Fine-tuning and Integration". *Proceedings of IEEE Spoken Language Technology Workshop (SLT)*, pp. 863-870.

73. Shih-Heng Wang[+], **Jiatong Shi**, Chien-yu Huang, Shinji Watanabe, and Hung-yi Lee. 2024. "Fusion of Discrete Representations and Self-Augmented Representations for Multilingual Automatic Speech Recognition". *Proceedings of IEEE Spoken Language Technology Workshop (SLT)*, pp. 247-254.

72. Yifeng Yu[+], **Jiatong Shi**, Yuning Wu[+], Shinji Watanabe. 2024. "VISinger2+: End-to-End Singing Voice Synthesis Augmented by Self-Supervised Learning Representation". *Proceedings of IEEE Spoken Language Technology Workshop (SLT)*, pp. 719-726.

71. Kristin Qi[+], **Jiatong Shi**, Caroline Summerour, John Batsis, and Xiaohui Liang. 2024. "Exploiting Longitudinal Speech Data via Voice Assistant Systems for Early Detection of Cognitive Decline". *Proceedings of IEEE International Conference on E-health Networking, Application & Services (HealthCom)*, pp. 1-6.

70. Ibrahim Said Ahmad, Antonios Anastasopoulos, Ondřej Bojar, Claudia Borg, Marine Carpuat, Roldano Cattoni, Mauro Cettolo, William Chen, Qianqian Dong, Marcello Federico, Barry Haddow, Dávid Javorský, Mateusz Krubiński, Tsz Kim Lam, Xutai Ma, Prashant Mathur, Evgeny Matusov, Chandresh Maurya, John McCrae, Kenton Murray, Satoshi Nakamura, Matteo Negri, Jan Niehues, Xing Niu, Atul Kr. Ojha, John Ortega, Sara Papi, Peter Polák, Adam Pospíšil, Pavel Pecina, Elizabeth Salesky, Nivedita Sethiya, Balaram Sarkar, **Jiatong Shi**, Claytone Sikasote, Matthias Sperber, Sebastian Stüker, Katsuhito Sudoh, Brian Thompson, Alex Waibel, Shinji Watanabe, Patrick Wilken, Petr Zemánek, and Rodolfo Zevallos. 2024. "Findings of the IWSLT 2024 Evaluation Campaign". *Proceedings of the 21st International Conference on Spoken Language Translation (IWSLT 2024)*. pp. 1-11.

69. Yuxun Tang[+], **Jiatong Shi**, Yuning Wu[+], and Qin Jin. 2024. "An Exploration on Singing MOS Prediction". *Proceedings of IEEE 14th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, pp. 651-655.

68. Yuning Wu[*+], Yifeng Yu[*+], **Jiatong Shi**, Tao Qian[+], Qin Jin. 2024. "A Systematic Exploration of Joint-training for Singing Voice Synthesis". *Proceedings of IEEE 14th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, pp. 289-293.

67. **Jiatong Shi**, Shih-Heng Wang[+], William Chen, Martijn Bartelds, Vanya Bannihatti Kumar, Jinchuan Tian, Xuankai Chang, Dan Jurafsky, Karen Livescu, Hung-yi Lee, and Shinji Watanabe. 2024. "ML-SUPERB 2.0: Benchmarking Multilingual Speech Models Across Modeling Constraints, Languages, and Datasets". *Proceedings of Interspeech*, pp. 1230-1234.

66. **Jiatong Shi**, Xutai Ma, Hirofumi Inaguma, Anna Sun, and Shinji Watanabe. 2024. "MMM: Multi-Layer Multi-Residual Multi-Stream Discrete Speech Representation from Self-supervised Learning Model". *Proceedings of Interspeech*, pp. 2569-2573.

65. **Jiatong Shi**[*], Yueqian Lin[*+], Xinyi Bai[+], Keyi Zhang, Yuning Wu[+], Yuxun Tang[+], Yifeng Yu[+], Qin Jin, and Shinji Watanabe. 2024. "Singing Voice Data Scaling-up: An Introduction to ACE-Opencpop and ACE-KiSing". *Proceedings of Interspeech*, pp. 1880-1884.

64. Yuning Wu[+], **Jiatong Shi**, Yifeng Yu[+], Yuxun Tang[+], Qian Tao[+], Yueqian Lin[+], Jionghao Han[+], Xinyi Bai[+], Shinji Watanabe, and Qin Jin. 2024. "Muskits-ESPnet: a Comprehensive Toolkit for Singing Voice Synthesis in New Paradigm". *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 11279-11281.

63. Kalvin Chang, Yi-Hui Chou[+], **Jiatong Shi**, Hsuan-Ming Chen, Nicole Holliday, Odette Scharenborg, and David R. Mortensen. 2024. "Self-supervised Speech Representations Still Struggle with African American Vernacular English". *Proceedings of Interspeech*, pp. 4643-4647.

62. Tejes Srivastava[+], **Jiatong Shi**, William Chen, and Shinji Watanabe. 2024. "EFFUSE: Efficient Self-Supervised Feature Fusion for E2E ASR in Multilingual and Low Resource Scenarios". *Proceedings of Interspeech*, pp. 3989-3993.

61. Jee-weon Jung[*], Wangyou Zhang[*], **Jiatong Shi**[*], Zakaria Aldeneh, Takuya Higuchi, Barry-John Theobald, Ahmed Hussen Abdelaziz, and Shinji Watanabe. 2024. "ESPnet-SPK: full pipeline speaker embedding toolkit with reproducible recipes, self-supervised front-ends, and off-the-shelf models". *Proceedings of Interspeech*, pp. 4278-4282.

60. Yifan Peng, Jinchuan Tian, William Chen, Siddhant Arora, Brian Yan, Yui Sudo, Muhammad Shakeel, Kwanghee Choi, **Jiatong Shi**, Xuankai Chang, Jee-weon Jung, and Shinji Watanabe. 2024. "OWSM v3.1: Better and Faster Open Whisper-Style Speech Models based on E-Branchformer". *Proceedings of Interspeech*, pp. 352-356.

59. Shuhua Li[+], Qirong Mao, and **Jiatong Shi**. 2024. "PL-TTS: A Generalizable Prompt-based Diffusion TTS Augmented by Large Language Model". *Proceedings of Interspeech*, pp. 4888-4892.

58. Xuankai Chang, **Jiatong Shi**, Jinchuan Tian, Yuning Wu[+], Yuxun Tang[+], Yihan Wu, Shinji Watanabe, Yossi Adi, Xie Chen, and Qin Jin. 2024. "The Interspeech 2024 Challenge on Speech Processing Using Discrete Units". *Proceedings of Interspeech*, pp. 2559-2563.

57. Yongyi Zang, **Jiatong Shi**, You Zhang, Ryuichi Yamamoto, Jionghao Han[+], Yuxun Tang[+], Shengyuan Xu, Wenxiao Zhao, Jing Guo, Tomoki Toda, Zhiyao Duan. 2024. "CtrSVDD: A Benchmark Dataset and Baseline Analysis for Controlled Singing Voice Deepfake Detection". *Proceedings of Interspeech*, pp. 4783-4787.

56. Yuxun Tang[+], Yuning Wu[+], **Jiatong Shi**, and Qin Jin. 2024. "SingOMD: Singing Oriented Multi-resolution Discrete Representation Construction from Speech Models". *Proceedings of Interspeech*, pp. 2564-2568.

55. Yuning Wu[+], Chunlei Zhang, **Jiatong Shi**, Yuxun Tang[+], and Qin Jin. 2024. "TokSing: Singing Voice Synthesis based on Discrete Tokens". *Proceedings of Interspeech*, pp. 2549-2553.

54. Taiqi He, Kwanghee Choi, Lindia Tjuatja, **Jiatong Shi**, Nate Robinson, Graham Neubig, Shinji Watanabe, David Mortensen, and Lori Levin. 2024. "WAV2GLOSS: Generating Interlinear Glossed Text from Speech". *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 568-582.

53. Rongjie Huang, Chunlei Zhang, Yongqi Wang, Dongchao Yang, Jinchuan Tian, Luping Liu, Zhenhui Ye, Ziyue Jiang, Xuankai Chang, **Jiatong Shi**, Chao Weng, Zhou Zhao, and Dong Yu. 2024. "Revisiting Voice Large Language Models as Scalable Multi-Lingual and Multi-Task Learners". *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)"*, pp. 10929-10942.

52. Dongchao Yang[*], Jinchuan Tian[*], Xu Tan, Rongjie Huang, Songxiang Liu, Xuankai Chang, **Jiatong Shi**, Sheng Zhao, Jiang Bian, Xixin Wu, Zhou Zhao, and Helen Meng. 2024. "UniAudio: An Audio Foundation Model Toward Universal Audio Generation". *Proceedings of Forty-Second International Conference on Machine Learning*.

51. Shu-wen Yang, Heng-Jui Chang[*], Zili Huang[*], Andy T. Liu[*], Cheng-I Lai[*], Haibin Wu[*], **Jiatong Shi**, Xuankai Chang, Hsiang-Sheng Tsai, Wen-Chin Huang, Tzu-hsun Feng, Po-Han Chi, Yist Y. Lin, Yung-Sung Chuang, Tzu-Hsien Huang, Wei-Cheng Tseng, Kushal Lakhotia, Abdelrahman Mohamed, Shang-Wen Li, Shinji Watanabe, and Hung-yi Lee. 2024. "A Large-Scale Evaluation of Speech Foundation Models". In *The IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 32: 2884-2899.

50. **Jiatong Shi**, Hirofumi Inaguma, Xutai Ma, Ilia Kulikov, and Anna Sun. 2024. "Multi-resolution HuBERT: Multi-resolution Speech Self-Supervised Learning with Masked Unit Prediction". *Proceedings of the Twelfth International Conference on Learning Representations (ICLR)*, SpotLight.

49. Rongjie Huang[*], Mingze Li[*], Dongchao Yang[*], **Jiatong Shi**[*], Xuankai Chang, Zhenhui Ye, Yuning Wu[+], Zhiqing Hong, Jiawei Huang, Jinglin Liu, Yi Ren, Zhou Zhao, Shinji Watanabe. 2024. "AudioGPT: Understanding and Generating Speech, Music, Sound, and Talking Head". *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 23802-23804.

48. Takashi Maekaku, **Jiatong Shi**, Xuankai Chang, Yuya Fujita, and Shinji Watanabe. 2024. "HuBERTopic: Enhancing semantic representation of HuBERT through self-supervision utilizing topic model". *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 11741-11745.

47. Chien-yu Huang, Ke-Han Lu, Shih-Heng Wang, Chi-Yuan Hsiao, Chun-Yi Kuan, Haibin Wu, Siddhant Arora, Kai-Wei Chang, **Jiatong Shi**, Yifan Peng, Roshan Sharma, Shinji Watanabe, Bhiksha Ramakrishnan, Shady Shehata, and Hung-yi Lee. 2024. "Dynamic-SUPERB: Towards A Dynamic, Collaborative, and Comprehensive Instruction-Tuning Benchmark for Speech". *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 12136-12140.

46. Xuankai Chang, Brian Yan, Kwanghee Choi, Jeeweon Jung, Yichen Lu, Soumi Maiti, Roshan Sharma, **Jiatong Shi**, Jinchuan Tian, Shinji Watanabe, Yuya Fujita, Takashi Maekaku, Pengcheng Guo, Yao-Fei Cheng, Pavel Denisov, Kohei Saijo, and Hsiu-Hsuan Wang. 2024. "Exploring Speech Recognition, Translation, and Understanding with Discrete Speech Units: A Comparative Study". *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 11481-11485.

45. Yi-Hui Chou[+], Kalvin Chang, Meng-Ju Wu, Winston Ou, Alice Wen-Hsin Bi, Carol Yang, Bryan Y. Chen, Rong-Wei Pai, Po-Yen Yeh, Jo-Peng Chiang, Lu-Tshiann Phoann, Winnie Chang, Chenxuan Cui, Noel Chen, and **Jiatong Shi**. 2023. "Evaluating Self-supervised Speech Models on a Taiwanese Hokkien Corpus". *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 1-8.

44. Wen-Chin Huang, Lester Phillip Violeta, Songxiang Liu, **Jiatong Shi**, and Tomoki Toda. 2023. "The Singing Voice Conversion Challenge 2023". *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 1-8.

43. **Jiatong Shi**, William Chen, Dan Berrebbi, Hsiu-Hsuan Wang, Wei-Ping Huang, En-Pei Hu, Ho-Lam Chuang, Xuankai Chang, Yuxun Tang[+], Shang-Wen Li, Abdelrahman Mohamed, Hung-yi Lee, and Shinji Watanabe. 2023. "Findings of the 2023 ML-SUPERB Challenge: Pre-Training and Evaluation over More Languages and Beyond". *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 1-8.

42. Yifan Peng, Jinchuan Tian, Brian Yan, Dan Berrebbi, Xuankai Chang, Xinjian Li, **Jiatong Shi**, Siddhant Arora, William Chen, Roshan Sharma, Wangyou Zhang, Yui Sudo, Muhammad Shakeel, Jee-Weon Jung, Soumi Maiti, and Shinji Watanabe. 2023. "Reproducing Whisper-Style Pre-training Using an Open-Source Toolkit and Publicly Available Data". *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 1-8.

41. William Chen, **Jiatong Shi**, Brian Yan, Dan Berrebbi, Wangyou Zhang, Yifan Peng, Xuankai Chang, Soumi Maiti, and Shinji Watanabe. 2023. "Joint Prediction and Denoising for Large-Scale Multilingual Self-Supervised Learning". *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 1-8.

40. Yui Sudo, Shakeel Muhhamad, Brian Yan, **Jiatong Shi**, and Shinji Watanabe. 2023. "4D: Joint Modeling of CTC, Attention, Transducer, and Mask-predict Decoders". *Proceedings of Interspeech*, pp. 3312-3316.

39. **Jiatong Shi**, Dan Berrebbi[*], William Chen[*], En-Pei Hu[*], Wei-Ping Huang[*], Ho Lam Chung[*], Xuankai Chang, Shang-Wen (Daniel) Li, Abdelrahman Mohamed, Hung-yi Lee, and Shinji Watanabe. 2023. "ML-SUPERB: Multilingual Speech Universal PERformance Benchmark". *Proceedings of Interspeech*, pp. 884-888.

38. **Jiatong Shi**, Yun Tang, Hirofumi Inaguma, Hongyu Gong, Juan Pino, and Shinji Watanabe. 2023. "Exploration on HuBERT with Multiple Resolution" *Proceedings of Interspeech*, pp. 3287-3291.

37. Sweta Agrawal, Antonios Anastasopoulos, Luisa Bentivogli, Ondřej Bojar, Claudia Borg, Marine Carpuat, Roldano Cattoni, Mauro Cettolo, Mingda Chen, William Chen, Khalid Choukri, Alexandra Chronopoulou, Anna Currey, Thierry Declerck, Qianqian Dong, Kevin Duh, Yannick Estève, Marcello Federico, Souhir Gahbiche, Barry Haddow, Benjamin Hsu, Phu Mon Htut, Hirofumi Inaguma, Dávid Javorský, John Judge, Yasumasa Kano, Tom Ko, Rishu Kumar, Pengwei Li, Xutai Ma, Prashant Mathur, Evgeny Matusov, Paul McNamee, John P. McCrae, Kenton Murray, Maria Nadejde, Satoshi Nakamura, Matteo Negri, Ha Nguyen, Jan Niehues, Xing Niu, Atul Kr. Ojha, John E. Ortega, Proyag Pal, Juan Pino, Lonneke van der Plas, Peter Polák, Elijah Rippeth, Elizabeth Salesky, **Jiatong Shi**, Matthias Sperber, Sebastian Stüker, Katsuhito Sudoh, Yun Tang, Brian Thompson, Kevin Tran, Marco Turchi, Alex Waibel, Mingxuan Wang, Shinji Watanabe, and Rodolfo Zevallos. 2023. "Findings of the IWSLT 2023 Evaluation Campaign", *Proceedings of the 20th International Conference on Spoken Language Translation (IWSLT)*, pp. 1-61

36. Brian Yan[*], **Jiatong Shi**[*], Soumi Maiti, William Chen, Xinjian Li, Yifan Peng, Siddhant Arora, and Shinji Watanabe. 2023. "CMU's IWSLT 2023 Simultaneous Speech Translation System", *Proceedings of the 20th International Conference on Spoken Language Translation (IWSLT)*, pp. 235-240

35. Biran Yan[*], **Jiatong Shi**[*], Yun Tang, Hirofumi Inaguma, Yifan Peng, Siddharth Dalmia, Peter Polák, Patrick Fernandes, Dan Berrebbi, Tomoki Hayashi, Xiaohui Zhang, Zhaoheng Ni, Moto Hira, Soumi Maiti, Juan Pino, Shinji Watanabe. 2023. "ESPnet-ST-v2: Multipurpose Spoken Language Translation Toolkit", *Proceedings of the 61th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 400–411

34. Tao Qian[+], Fan Lou, **Jiatong Shi**, Yuning Wu[+], Shuai Guo[+], Xiang Yin, and Qin Jin. 2023. "UniLG: A Unified Structure-aware Framework for Lyrics Generation", *Proceedings of the 61th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 983–1001.

33. William Chen, Brian Yan, **Jiatong Shi**, Yifan Peng, Soumi Maiti, and Shinji Watanabe. 2023. "Improving Massively Multilingual ASR with Auxiliary CTC Objectives" *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1-5.

32. **Jiatong Shi**, Chan-Jan Hsu, Holam Chung, Dongji Gao, Paola Garcia, Shinji Watanabe, Ann Lee, and Hung-yi Lee. 2023. "Bridging Speech and Text Pre-trained Models with Unsupervised ASR" *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1-5.

31. **Jiatong Shi**, Yun Tang, Ann Lee, Hirofumi Inaguma, Changhan Wang, Juan Pino, and Shinji Watanabe. 2023. "Enhancing Speech-to-Speech Translation with Multiple TTS Targets" *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1-5.

30. Dongji Gao[*], **Jiatong Shi**[*], Shun-Po Chuang, Leibny Paola Garcia, Hung-yi Lee, Shinji Watanabe, and Sanjeev Khudanpur. 2023. "EURO: ESPnet Unsupervised ASR Open-source Toolkit" *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1-5.

29. Yuning Wu[+], **Jiatong Shi**, Tao Qian[+], and Qin Jin. 2023. ”PHONEix: Acoustic Feature Processing Strategy for Enhanced Singing Pronunciation with Phoneme Distribution Predictor” *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1-5.

28. Antonios Anastasopoulos, Loïc Barrault, Luisa Bentivogli, Marcely Zanon Boito, Ondřej Bojar, Roldano Cattoni, Anna Currey, Georgiana Dinu, Kevin Duh, Maha Elbayad, Clara Emmanuel, Yannick Estève, Marcello Federico, Christian Federmann, Souhir Gahbiche, Hongyu Gong, Roman Grundkiewicz, Barry Haddow, Benjamin Hsu, Dávid Javorský, Věra Kloudová, Surafel Lakew, Xutai Ma, Prashant Mathur, Paul McNamee, Kenton Murray, Maria Nădejde, Satoshi Nakamura, Matteo Negri, Jan Niehues, Xing Niu, John Ortega, Juan Pino, Elizabeth Salesky, **Jiatong Shi**, Matthias Sperber, Sebastian Stüker, Katsuhito Sudoh, Marco Turchi, Yogesh Virkar, Alexander Waibel, Changhan Wang, and Shinji Watanabe. 2022. ”Findings of the IWSLT 2022 Evaluation Campaign”, *Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT)*, pp. 298-307

27. Yen Meng, Hsuan-Jui Chen, **Jiatong Shi**, Shinji Watanabe, Paola Garcia, Hung-yi Lee, and Hao Tang. 2022. ”On Compressing Sequences for Self-Supervised Speech Models” *IEEE Spoken Language Technology Workshop (SLT)*, pp. 1128-1135.

26. Tzu-hsun Feng, Annie Dong, Ching-Feng Yeh, Shu-wen Yang, Tzu-Quan Lin, **Jiatong Shi**, Kai-Wei Chang, Zili Huang, Haibin Wu, Xuankai Chang, Shinji Watanabe, Abdel-rahman Hohamed, Shang-Wen Li, and Hung-yi Lee. 2022. ”SUPERB @ SLT 2022: Challenge on Generalization and Efficiency of Self-Supervised Speech Representation Learning” *Proceedings of IEEE Spoken Language Technology Workshop (SLT)*, pp. 1096-1103.

25. **Jiatong Shi**, George Saon, David Haws, Shinji Watanabe and Brian Kingsbury. 2022. ”VQ-T: RNN Transducers using Vector-Quantized Prediction Network States” *Proceedings of Interspeech*, pp. 1656-1660.

24. **Jiatong Shi**, Shuai Guo[+], Tao Qian[+], Nan Huo, Tomoki Hayashi, Yuning Wu[+], Fangzheng Xu, Xuankai Chang, Huazhe Li, Peter Wu, Shinji Watanabe and Qin Jin. 2022. ”Muskits: an End-to-end Music Processing Toolkit for Singing Voice Synthesis ” *Proceedings of Interspeech*, pp.4277-4281.

23. Shuai Guo[*+], **Jiatong Shi**[*], Tao Qian[+], Shinji Watanabe and Qin Jin. 2022. ”SingAug: Data Augmentation for Singing Voice Synthesis with Cycle-consistent Training Strategy ” *Proceedings of Interspeech*, pp.4272-4276.

22. Dan Berrebbi[+], **Jiatong Shi**, Brian Yan, Osbel López-Francisco, Jonathan Amith and Shinji Watanabe. 2022. ”Combining Spectral and Self-Supervised Features for Low Resource Speech Recognition and Translation” *Proceedings of Interspeech*, pp.3533-3537.

21. Keqi Deng, Shinji Watanabe, **Jiatong Shi** and Siddhant Arora. 2022. ”Blockwise Streaming Transformer for Spoken Language Understanding and Simultaneous Speech Translation ” *Proceedings of Interspeech*, pp.1746-1750.

20. Yui Sudo, Shakeel Muhammad, Kazuhiro Nakadai, **Jiatong Shi** and Shinji Watanabe. 2022. ”Streaming Automatic Speech Recognition with Re-blocking Processing Based on Integrated Voice Activity Detection” *Proceedings of Interspeech*, pp.4641-4645.

19. Brian Yan, Patrick Fernandes, Siddharth Dalmia, **Jiatong Shi**, Yifan Peng, Dan Berrebbi, Xinyi Wang, Graham Neubig, and Shinji Watanabe. 2022. “CMU’s IWSLT 2022 Dialect Speech Translation System”, *Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT)*, pp. 298-307

18. Hsiang-Sheng Tsai, Heng-Jui Chang, Wen-Chin Huang, Zili Huang, Kushal Lakhotia, Shu-wen Yang, Shuyan Dong, Andy T. Liu, Cheng-I Lai, **Jiatong Shi**, Xuankai Chang, Phil Hall, Shang-Wen Li, Shinji Watanabe, Abdelrahman Mohamed, and Hung-yi Lee. 2022. ”SUPERB-SG: Enhanced Speech processing Universal PERformance Benchmark for Semantic and Generative Capabilities” *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 8479-8492.

17. Tao Qian[+], **Jiatong Shi**, Shuai Guo[+], Peter Wu[+], and Qin Jin. 2022. ”Training strategies for automatic song writing: a perspective with a unified framework” *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4738-4742.

16. Chunlei Zhang, **Jiatong Shi**, Chao Weng, Meng Yu, and Dong Yu. 2022. ”Towards End-to-end Speaker Diarization with Generalized Neural Speaker Clustering” *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8372-8376.

15. **Jiatong Shi**, Chunlei Zhang, Chao Weng, Shinji Watanabe, Meng Yu, and Dong Yu. 2022. ”An Investigation of Neural Uncertainty Estimation for Target Speaker Extraction Equipped RNN Transducer”. *Computer Speech and Language (CSL)*. 73: 10327.

14. Peter Wu[+], **Jiatong Shi**, Yifan Zhong, Shinji Watanabe, and Alan W Black. 2021. ”Acoustic Cross-lingual Transfer using Language Similarity”. *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 1050-1057.

13. Peter Wu[+], Paul Pu Liang, **Jiatong Shi**, Ruslan Salakhutdinov, and Shinji Watanabe, Louis-Philippe Morency. 2021. "Understanding the Tradeoffs in Client-side Privacy for Downstream Speech Tasks". *Proceedings of APSIPA Annual Summit and Conference*, pp. 841-848.

12. Shu-wen Yang, Po-Han Chi[*], Yung-Sung Chuang[*], Cheng-I Jeff Lai[*], Kushal Lakhotia[*], Yist Y Lin[*], Andy T Liu[*], **Jiatong Shi**[*], Xuankai Chang, Guan-Ting Lin, Tzu-Hsien Huang, Wei-Cheng Tseng, Ko-tik Lee, Da-Rong Liu, Zili Huang, Shuyan Dong, Shang-Wen Li, Shinji Watanabe, Abdelrahman Mohamed, and Hung-yi Lee. 2021. "SUPERB: Speech processing Universal PERformance Benchmark". *Proceedings of Interspeech*, pp. 1194-1198.

11. Hirofumi Inaguma, Brian Yan, Siddharth Dalmia, Pengcheng Guo, **Jiatong Shi**, Kevin Duh, and Shinji Watanabe. 2021. "ESPnet-ST IWSLT 2021 Offline Speech Translation System" *Proceedings of the 18th International Conference on Spoken Language Translation (IWSLT)*, pp.100-109.

10. **Jiatong Shi**, Jonathan D. Amith, Xuankai Chang, Siddharth Dalmia, Brian Yan, and Shinji Watanabe. 2021. "Highland Puebla Nahuatl–Spanish Speech Translation Corpus for Endangered Language Documentation". *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, pp.53-63.

9. Jonathan D. Amith, **Jiatong Shi**, and Rey Castillo García. 2021. "End-to-End Automatic Speech Recognition: Its Impact on the Workflow for Documenting Yoloxóchitl Mixtec". *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, pp. 64-80.

8. **Jiatong Shi**, Chunlei Zhang, Chao Weng, Shinji Watanabe, Meng Yu, and Dong Yu. 2021. "Improving RNN Transducer with Target Speaker Extraction and Neural Uncertainty Estimation". *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6908-6912.

7. **Jiatong Shi**[*], Shuai Guo[*+], Nan Huo, Yuekai Zhang, and Qin Jin. 2021. "Sequence-to-sequence Singing Voice Synthesis with Perceptual Entropy Loss". *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 76-80.

6. Pengcheng Guo, Florian Boyer, Xuankai Chang, Tomoki Hayashi, Yosuke Higuchi, Hirofumi Inaguma, Naoyuki Kamo, Chenda Li, Daniel Garcia-Romero, **Jiatong Shi**, Jing Shi, Shinji Watanabe Kun Wei, Wangyou Zhang, and Yuekai Zhang. 2021. "Recent Developments on ESPNet Toolkit Boosted by Conformer". *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.5874-5878.

5. **Jiatong Shi**, Jonathan D. Amith, Rey Castillo García, Esteban Guadalupe Sierra, Kevin Duh, and Shinji Watanabe. 2021. "Leveraging End-to-End ASR for Endangered Language Documentation: An Empirical Study on Yolóxochitl Mixtec". *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pp.1134-1145.

4. **Jiatong Shi**, Kunlin Yang, Wei Xu, and Mingming Wang. 2021. "Leveraging deep learning with audio analytics to predict the success of crowdfunding projects." *Journal of Supercomputing*. 77(7): 7833-7853.

3. **Jiatong Shi**, Nan Huo, and Qin Jin. 2020. "Context-aware Goodness of Pronunciation for Computer Assisted Pronunciation Training". *Proceedings of Interspeech*, pp. 3057-3061.

2. Wenxin Hou, Yue Dong, Bairong Zhuang, Longfei Yang, **Jiatong Shi**, and Takahiro Shinozaki. 2020. "Large-Scale End-to-End Multilingual Speech Recognition and Language Identification with Multi-Task Learning". *Proceedings of Interspeech*, pp. 1047-1051.

1. **Jiatong Shi**, Wei Du, and Wei Xu. 2018. "Identifying Impact Factors of Question Quality in Online Health Q&A Communities: an Empirical Analysis on MedHelp." *Proceedings of the 22nd Pacific Asia Conference on Information Systems (PACIS)*, Association of Information Systems, pp. 1886- 1898.

## Arxiv Preprint

6. Shih-Heng Wang[+], Zih-Ching Chen, **Jiatong Shi**, Ming-To Chuang, Guan-Ting Lin, Kuan-Po Huang, David Harwath, Shang-Wen Li, and Hung-yi Lee. "How to Learn a New Language? An Efficient Solution for Self-Supervised Learning Models Unseen Languages Adaptation in Low-Resource Scenario".

5. You Zhang, Yongyi Zang, **Jiatong Shi**, Ryuichi Yamamoto, Jionghao Han[+], Yuxun Tang[+], Tomoki Toda, and Zhiyao Duan. "SVDD Challenge 2024: A Singing Voice Deepfake Detection Challenge Evaluation Plan".

4. Yuxun Tang[+], **Jiatong Shi**, Yuning Wu[+], and Qin Jin. "SingMOS: An extensive Open-Source Singing Voice Dataset for MOS Prediction"

3. Yu-Kuan Fu, Liang-Hsuan Tseng, **Jiatong Shi**, Hung-yi Lee, and Shinji Watanabe. "Improving Speaker-preserving Cascaded Unsupervised Speech Translation with Denoising Back-translation"

2. Yushi Hu, Chunlei Zhang, **Jiatong Shi**, Jiachen Lian, Mari Ostendorf, Dong Yu. ”ProsodyBERT: Self-Supervised Prosody Representation for Style-Controllable TTS”

1. Tomoki Hayashi, Ryuichi Yamamoto, Takenori Yoshimura, Peter Wu[+], **Jiatong Shi**, Takaaki Saeki, Yooncheol Ju, Yusuke Yasuda, Shinnosuke Takamichi, Shinji Watanabe. ”ESPnet2-TTS: Extending the Edge of TTS Research”

### In Review

**Jiatong Shi**, Jionghao Han[+], Yichen Lu, Santiago Pascual, Chenye Cui, Pengfei Wu, Shinji Watanabe, Chao Weng, and Cong Zhou. ”Speech-DRAME: Speech Detailed Role-play Assessment with Modeling and Evaluation”

Jionghao Han[+], **Jiatong Shi**, Yuxun Tang[+], Yiwen Zhao[+], and Shinji Watanabe. ”CartoonSVS: Beyond Human Timbres in Singing Voice Synthesis”

Jionghao Han[+], **Jiatong Shi**, Masao Someki, Yuxun Tang[+], Lan Liu[+], Yiwen Zhao[+], Wenhao Feng[+], and Shinji Watanabe. ”SingingSDS: A Singing-Capable Spoken Dialogue System for Conversational Roleplay Applications”

Yuxun Tang[+], Lan Liu[+], **Jiatong Shi**, Jionghao Han[+], Yiwen Zhao[+], and Qin Jin. ”SingMOS-Pro: A Comprehensive MOS Dataset for Singing Voice Synthesis”

Bo-Hao Su, Hui-Ying Shih, Jinchuan Tian, **Jiatong Shi**, Chi-Chun Lee, Carlos Busso, and Shinji Watanabe. ”Reasoning Beyond Majority Vote: An Explainable SpeechLM Framework for Speech Emotion Recognition”

Lester Phillip Violeta, Xueyao Zhang, **Jiatong Shi**, Yusuke Yasuda, Wen-Chin Huang, Zhizheng Wu, and Tomoki Toda. ”The Singing Voice Conversion Challenge 2025: Beyond Voice Identity toward Singing Style Control”

Siddhant Arora, Jinchuan Tian, Hayato Futami, **Jiatong Shi**, Yosuke Kashiwagi, Emiru Tsunoo, and Shinji Watanabe. ”Chain-of-Thought Reasoning in Streaming Full-Duplex End-to-End Spoken Dialogue Systems”

Jinchuan Tian, Bo-Hao Su, Siddhant Arora, **Jiatong Shi**, William Chen, Keita Goto, Takashi Maekaku, Yusuke Shinohara, Chao-Han Huck Yang, and Shinji Watanabe. ”Opus-Chat: Exploring How Speech Language Models Remember and Think”

## Awards, Fellowships, Grants, & Notable Achievement

2024 **Best Paper Award**, ISCA Interspeech
**Best Paper Award**, EMNLP
**Honorable Mention Demo Award**, ACMMM
**Best Paper Honorable Mention**, Responsible Speech Foundation Models Workshop (Special Session at ISCA Interspeech)
**1st Place**, Discrete Speech Challenge (SVS Track)
**2nd Place**, Discrete Speech Challenge (ASR Track)

2023 **Best Paper Finalist at 2023 ASRU**, IEEE ASRU
**Best Paper Finalist at 2022 SLT**, IEEE SLT
**Best Project at 2022 SLT Hackathon (as a mentor)**, IEEE SLT

2022 **CMU Presidential Fellowship**, Carnegie Mellon University
**1st Place**, IWSLT 2022 (the shared task in dialectal speech translation track)
**7th Place (Finalist)**, AI Song Contest 2022

2021 **Research Fellowship**, Language Technologies Institute, Carnegie Mellon University

2019 **Outstanding Graduates**, Renmin University of China
**Undergraduate Research Grants**, Renmin University of China
**Undergraduate Grants**, in National Training Programs of Innovation and Entrepreneurship for Undergraduates

2018 **Award of Academic Excellence**, School of Information, Renmin University of China
**Special Award**, in National University Data-driven Innovation & Research Contest, Peking University

2017 **Award of Academic Excellence**, School of Information, Renmin University of China

2016 **Award of Academic Excellence**, School of Information, Renmin University of China

## Presentations

Sep. 2025. *Automatic Quality Assessment for Speech and Beyond*. Invited Talk, Speech Lunch at CMU.

Aug. 2025. *Automatic Quality Assessment for Speech and Beyond*. Tutorial, together with Dr. Wen-Chin Huang and Dr. Erica Cooper, Interspeech2025.

Jul. 2025. *Multi-metric Speech Evaluation*. Invited Talk, Inflection AI.

Mar. 2025. *ESPnet-Codec*. Invited Talk, Amazon at Pittsburgh.

Mar. 2025. *Neural Codec*. Lecture, Speech Processing for Conversational AI, CMU.

Mar. 2025. *Improving Universality for Speech Representation Learning*. Invited Talk, Anuttacon.

Feb. 2025. *Singing Voice Synthesis: Data curation, Modeling, and Evaluation*. Invited Talk, Conversational AI Reading Group at Mila.

Jan. 2025. *Improving Universality for Speech Representation Learning*. Ph.D. Thesis Proposal at CMU.

Oct. 2024. *ESPnet-Codec and VERSA*. Invited Talk, Speech Lunch (Sphinx Lunch) at CMU.

Aug. 2024. *The Revolution of AI-driven Speech Communication: a Discussion on the Technical Roadmap behind GPT-4o*. Invited Talk, Exodus (Podcast online in Chinese).

Mar. 2024. *Multi-resolution HuBERT*. Invited Talk, Algorithm Talk at Alibaba-NTU (Nanyang Technological University).

Mar. 2024. *Speech Representation Learning: From Continuous to Discrete*. Invited Talk, Huawei.

Mar. 2024. *Speech-to-speech Translation: Task Overview*. Lecture, Speech Processing, CMU.

Dec. 2023. *ML-SUPERB and ML-SUPERB Challenge*. Invited Talk, SPARKS Workshop at ASRU 2023, National Taiwan University.

Aug. 2023. *Language Resources and Technology: Inclusion, Diversity, and Sovereignty*. Panel Discussion, SIGUL Workshop, Interspeech@Dublin.

Jun. 2023. *ESPnet-Muskits and Its Integration to LLM*. Invited Talk, Music & Audio Processing Workshop, HKUST.

Apr. 2023. *Speech-to-speech Translation*. Lecture, Speech Processing, CMU.

Feb. 2023. *ESPnet Tutorial 2: New Tasks and Models*. Lecture, Speech Processing, CMU.

Nov. 2022. *Progress in Singing Voice Synthesis with Muskits*. Invited Talk, AIR lab, University of Rochester.

Sep. 2022. *End-to-End Unsupervised ASR and Its Application*. Invited Talk, Sphinx Lunch at CMU.

Sep. 2022. *ESPnet Tutorial 2: New Tasks and Models*. Lecture, Speech Recognition and Understanding, CMU.

Jun. 2022. *ESPnet: an End-to-end Speech Processing Toolkit*. Invited Talk, 2022 JSALT Summer Workshop at JHU.

Apr. 2022. *Speech Technologies with Neural Networks*. Guest Lecture, Natural Language Processing, CMU.

## Teaching Experience

| | | |
|---|---|---|
| Spring 2025 | **Speech Processing for Conversational AI (CS492/692)**, Teaching Assistant | *CMU* |
| Spring 2024 | **Speech Processing for Conversational AI (CS492/692)**, Teaching Assistant | *CMU* |
| Spring 2023 | **Speech Processing (CS492/692)**, Teaching Assistant | *CMU* |
| Fall 2022 | **Speech Recognition and Understanding (CS751)**, Teaching Assistant | *CMU* |
| Spring 2022 | **Natural Language Processing (CS411/611/711)**, Teaching Assistant | *CMU* |
| Spring 2021 | **Principles of Programming Language (CS425/625)**, Course Assistant | *JHU* |
| Fall 2020 | **Introduction to Human Language Technologies (CS465/665)**, Course Assistant | *JHU* |
| Spring 2020 | **Machine Learning (CS475/675)**, Course Assistant | *JHU* |

## Students Mentored

2025- **Emily Isbell**, Briarcliff High School, Briarcliff Science Research Program
2025- **Zhuoyan (Terry) Tao**, Undergrad at USC
2025- **Haoran Wang**, Undergrad at SJTU
2025- **Wenhao Feng**, MCS at RUC
2024- **Mingda Liu**, PostDoc at Tongji University
2024- **Yiwen Zhao**, MSCV at CMU
2023- **Jionghao Han**, MS ECE at CMU
2022- **Yuxun Tang**, MCS at RUC
2025 **Lan Liu**, Undergrad at RUC
2022-2024 **Yifeng Yu**, MSMT at Georgia Tech
2023-2024 **Shin-Heng (Stan) Wang**, PhD student at USC (was MCS at NTU)
2023-2024 **Kristin Qi**, PhD student at UMass-Boston
2022-2024 **Tejes Srivastava**, MCS at UChicago (was Undergrad at UC Davis, REUSE Internship)
2024 **Shuhua Li**, MCS at Jiangsu University
2022-2024 **Yuning Wu**, Alibaba (was MCS at RUC)
2022-2023 **Yi-Hui (Sophia) Chou**, Zoom (was MIIS at CMU)
2023 **Yueqian Lin**, PhD student at DKU (was Undergraduate at DKU)
2023 **Xinyi Bai**, Google/Cornell
2021-2023 **Tao Qian**, Shanghai High School (was MCS at RUC)
2022-2023 **Tanisha Saxena**, Undergraduate at CMU (AI mentor program)
2021-2022 **Dan Berrebbi**, Apple (was MLT at CMU)
2019-2022 **Shuai Guo**, Alibaba (was MCS at RUC)
2021 **Peter Wu**, PhD student at UC Berkeley (was MSML at CMU)

## Professional Activities

### Peer Reviewer

- Journals:
  - IEEE Signal Processing Letters
  - Journal of Selected Topics in Signal Processing
  - IEEE Transactions on Audio, Speech, and Language Processing
  - IEEE Transactions on Neural Networks and Learning Systems
  - Computer Speech and Language
  - Language Resources and Evaluation
  - APSIPA Transactions on Signal and Information Processing
  - Transactions on Multimedia Computing, Communications and Applications
  - Speech Communication
  - KSII Transactions on Internet and Information Systems
  - Journal of Human-Machine Research
  - Signal, Image and Video Processing
  - Journal of Medical Internet Research
- Conferences:
  - Interspeech
  - ACL Rolling Review (ARR) and related CL conferences (ACL, NAACL, EMNLP)
  - ACM Multimedia
  - IEEE SLT
  - AAAI

  - ICASSP
  - ASRU
  - LREC
  - CVPR
  - ICLR
  - NeurIPS

## Open-source Contribution

- Toolkit:
  - ESPnet (https://github.com/espnet/espnet)
  - VERSA (https://github.com/shinjiwlab/versa)
  - S3PRL (https://github.com/s3prl/s3prl)
  - Muskits (Merged to ESPnet, archived url at https://github.com/SJTMusicTeam/Muskits)
  - ParallelWaveGAN (https://github.com/kan-bayashi/ParallelWaveGAN)
  - AudioGPT (https://github.com/AIGC-Audio/AudioGPT)
  - Fairseq (https://github.com/facebookresearch/fairseq)
- Dataset:
  - SingMOS Dataset (https://huggingface.co/datasets/TangRain/SingMOS)
  - SVDD Challenge Dataset (https://zenodo.org/records/10467648)
  - ACE-Opencpop (https://huggingface.co/datasets/espnet/ace-opencpop-segments)
  - ACE-KiSing / KiSing-V2 (https://huggingface.co/datasets/espnet/kising-v2-segments)
  - Highland Puebla Nahuatl (http://www.openslr.org/92/)
  - Yoloxóchitl Mixtec (http://www.openslr.org/89/)
  - Totonac (http://www.openslr.org/107/)
  - KiSing (http://shijt.site/index.php/2021/05/16/)
  - ML-SUPERB (https://arxiv.org/abs/2305.10615)

## Organizing Services

- Benchmark and Challenges:
  - SVCC (Singing voice Conversion Challenge), Expected AAAI 2025
  - Singing Voice Deepfake Detection Challenge (SVDD) at MIREX, ISMIR 2024
  - ML-SUPERB 2.0 at Interspeech 2024
  - Singing Voice Deepfake Detection Challenge (SVDD) at SLT 2024
  - Speech Processing Using Discrete Speech Unit Challenge at Interspeech 2024
  - ML-SUPERB (Multilingual SUPERB) at ASRU 2023
  - SVCC (Singing Voice Conversion Challenge) at ASRU 2023
  - IWSLT 2022 and 2023 (Simultaneous Track)
  - SUPERB
- Workshops:
  - SPARKS Workshop